The Choice of Prior in Bayesian Modeling of the Information Sampling Task

Axelsen, Martin Christian; Jepsen, Jens Richardt Møllegaard; Bak, Nikolaj

Published in:
Biological Psychiatry

Link to article, DOI:
10.1016/j.biopsych.2017.04.021

Publication date:
2018

Document Version
Early version, also known as pre-print

Citation (APA):
Axelsen, M. C., Jepsen, J. R. M., & Bak, N. (2018). The Choice of Prior in Bayesian Modeling of the Information Sampling Task. *Biological Psychiatry*, *83*(12), E59-E60. https://doi.org/10.1016/j.biopsych.2017.04.021

TITLE:

The choice of prior in Bayesian modeling of the Information Sampling Task

LIST OF AUTHORS and AFFILIATIONS:

Martin Christian Axelsen[1,2,3], Jens Richardt Møllegaard Jepsen[2,3,4], Nikolaj Bak[2,3]

1. Cognitive Systems - DTU Compute, Technical University of Denmark, Kgs. Lyngby, Denmark
2. Lundbeck Foundation Center for Clinical Intervention and Neuropsychiatric Schizophrenia Research (CINS), Mental Health Services Glostrup, Capital Region, Denmark
3. Center for Neuropsychiatric Schizophrenia Research (CNSR), Mental Health Services Glostrup, Capital Region, Denmark
4. Child and Adolescent Mental Health Centre, Mental Health Services in the Capital Region of Denmark, Copenhagen, Denmark.

CORRESPONDING AUTHOR

Martin Christan Axelsen
Technical University of Denmark
Richard Petersens Plads
Building 321, room 106
2800 Kgs. Lyngby
Phone: +45 45253922
e-mail: maxe@dtu.dk

Word count in main text: 1000

Number of figures and tables: 0

Equations: 4

To the editor:

The Information Sampling Task (IST) was introduced in Clark et al. (1) as a measure of reflection impulsivity. Performance on the IST is evaluated through the average number of boxes opened for each trial of the test, whether the chosen color was in majority or not, and ”the probability of the subject being correct at the point of decision [*P(correct)*]” (1, p 517). *P(correct)* is proposed calculated as a per-trial probability:

$$P(correct) = \frac{\sum_{k=A}^{Z} \binom{Z}{k}}{2^Z} \qquad (1)$$

If n_1 denotes the number of boxes opened of the chosen color and n_2, the number of boxes opened of the unchosen color, then $Z = 25 - (n_1 + n_2)$ is the number of unopened boxes, and $A = 13 - n_1$ is the additional number of boxes required for a majority. This formula is the cumulative distribution function over the probabilities that the choice of color is correct. Where the probability of seeing the chosen color in a box is described by the binomial distribution with probability 0.5.

In the recent correspondence by Bennett et al. (2) it was suggested that the original equation (eq. 1) does not encompass the correct dependency structures in the problem. As a correction, it was suggested that the problem is formulated in a Bayesian manner, with the hypergeometric distribution (3) as the likelihood and a uniform prior.

Bayes' theorem relates the 'posterior' probability to the 'likelihood', 'prior', and the 'model evidence' (see e.g. (4, eq. 1.12)):

$$posterior = \frac{likelihood \times prior}{model\ evidence}$$

The posterior probability is the probability of seeing a specific output given the observations e.g. that the majority color is red given some observations.

The likelihood is the probability of seeing the observations given the outcome or parameters e.g. the probability of seeing the colors that are seen, given that the majority color is red.

The prior incorporates knowledge about the outcome or parameters that exist before observing anything e.g. the prior probability that the color in a box is red.

The model evidence is the probability of observing the dataset. It functions as a normalizing coefficient, making the posterior a proper mathematical distribution.

The main advantage in using Bayesian modeling is in the possibility of more explicitly modeling assumptions and uncertainties in a problem. The Bayesian formulation of the *P(correct)* equation is based on the assumptions that are put in the likelihood and the prior. These quantities must therefore be chosen carefully. The choice of the hypergeometric distribution as likelihood seems fit.

When it comes to the IST, we are encountered with a prior probability used by the software to distribute the colors in the boxes. We call this the 'true prior' and it is based on how the test is coded in the software. We have contacted Dr. John Evenden, Director of science at Cambridge Cognition and Dr. Luke Clark, first author of (1) to ask what the true distribution is from which the different trials are created. The test paradigm was created with a fixed set of trials created from a binomial distribution where ratios more extreme than 20:5 were excluded. Furthermore the trials were created such that they would be similar across the two different conditions of the IST.

Due to the handpicking of trials, the true prior is not given by an exact mathematical distribution, and we therefore need to approximate it. Considering a binomial distribution, which was used to create the initial set of trials, the chance of randomly obtaining one or more trials with a ratio more extreme than 20:5 when sampling 21 times (since there are 21 trials in the IST) is approximately 0.02. Given the constraints of the IST, we therefore believe that the binomial distribution is a reasonable estimate of the true prior.

The distribution with which the colors are distributed is however not revealed to the subjects prior to the first trial, and they will therefore initiate the IST with a prior assumption of their own about how the colors are distributed - we will denote this prior the 'personal prior'. As the IST consists of multiple trials, and the pattern of colors is revealed after each trial, the subjects are likely to adjust their assumption (personal prior) over the duration of the test, such that they assume that extreme cases are unlikely. The personal prior will therefore approach the binomial distribution and not a uniform distribution.

We believe that it is reasonable to calculate *P(correct)* with a Bayesian equation, though we believe that the prior should be binomial and not uniform as suggested in Bennett et al. (2). The Bayesian model using the binomial distribution, where the probability of each class is 0.5 as prior, and with the hypergeometric distribution as likelihood, is given below (3):

$$P(\theta|n_1, n_2) = \frac{\binom{\theta}{n_1}\binom{N-\theta}{n_2}\binom{N}{\theta}}{\sum_{j=n_1}^{N-n_2}\binom{j}{n_1}\binom{N-j}{n_2}\binom{N}{j}}$$

$$P(correct) = P(\theta \geq 13|n_1, n_2) = \sum_{M=13}^{25} P(\theta = M|n_1, n_2)$$

Where *N* is the total number of boxes – in this case 25 and θ is the number of boxes containing the chosen color. Given the relatively simple constraints, the Bayesian formulation of this statistic can be shown to be exactly equal to the original equation from (1) (eq. 1).

As shown above, the original equation formulated by Clark et al. (1) includes the relevant information regarding opened boxes, and models the correct dependencies of the problem. We therefore suggest that for calculating *P(correct)* for the IST, this equation should still be used.

As mentioned above, the personal prior can differ from the true prior in the beginning of the task, which indicates that there will be a learning period. This period is likely different from subject to subject. The *P(correct)* measure is hence a more complex construct and probably reflects several cognitive functions. If a more specific outcome measure is desirable we suggest that the test subjects are introduced to the prior in the test instructions. The following sentence could be added to the instruction: ”The color in each box is determined by the computer as *heads or tails* before the game starts”. Most -if not all subjects will then have an idea of the distribution, even if they are not familiar with the technicalities of the binomial distribution.

Acknowledgments

This work was supported by the Lundbeck Foundation (R155-2013-16337). We thank Dr. Luke Clark and Dr. John Evenden for assistance with understanding the construction of the IST. The authors report no financial disclosures or conflicts.

References

1. Clark L, Robbins TW, Ersche KD, Sahakian BJ (2006): Reflection Impulsivity in Current and Former Substance Users. *Biol Psychiatry*. 60: 515–522.
2. Bennett D, Oldham S, Dawson A, Parkes L, Murawski C, Yücel M (2016): Systematic Overestimation of Reflection Impulsivity in the Information Sampling Task. *Biol Psychiatry*.

3. Link WA, Barker RJ (2009): Bayesian Inference: with ecological applications. 1st ed. Academic Press, pp 110–115.

4. Bishop CM (2006): *Pattern recognition and Machine Learning*. Springer.